Non-context-free Languages

How do we show something is not context free?

The Pumping Lemma

Theorem. **(Pumping Lemma) Let L be context free. There exists a positive integer p with the following property.**

For every $w \in L$ of length at least p, w is divided into five parts, u, v, x, y, z, such that

- $|vy| \geq 1$,
- $|vxy| \leq p$, and
- for each $i \geq 0$, $uv^i xy^i z \in L$.

The Pumping Lemma

Theorem. **(Pumping Lemma) Let L be context free. There exists a positive integer p with the following property.**

For every $w \in L$ of length at least p, w is divided into five parts, u, v, x, y, z, such that

- $|vy| \geq 1$,
- $|vxy| \leq p$, **and**
- **for each** $i \geq 0$, $uv^ixy^iz \in L$.

The differences between this pumping lemma and the previous one.

- **There are two components that are jointly inserted or deleted.**
- **The part vxy may not be at the beginning of w.**

Proving the Pumping Lemma

Let $L = L(G)$ for some CNF grammar $G = (V, \Sigma, R, S)$.

If L is finite (i.e., has only a finite number of members), then there is a length k such that each member of L has length less than k. We have only to choose p to be k.

So we will assume L is infinite.

Proving the Pumping Lemma

Set $m = \|V\|$ and $p = 2^m$.

Let w be an arbitrary member of L having length at least p. Let T be a derivation tree for w.

Since G is a CNF grammar, for each subtree of T, the following properties hold:

- Each non-leaf node of R is a variable.
- Each leaf of R is a terminal.
- Each leaf of R is a unique child of its parent.
- Except for the leaves and their parents each node of R has exactly two children.
- The concatenation of the leaves of R is a substring of w.

A Useful Property

An **ancestor–descendant pair with identical label (ADPIL, for short)** in a production tree R is a node pair (r, s) such that

- r is an ancestor of s and
- the label of r is identical to the label of s (and thus, the label is a nonterminal).

Proof (cont'd)

Claim. **If R has more than $p/2 = 2^{m-1}$ leaves, then R contains an ADPIL.**

Proof (cont'd)

Claim. If R has more than $p/2 = 2^{m-1}$ leaves, then R contains an ADPIL.

Proof

Suppose R is a subtree of T with at least $2^{m-1} + 1$ leaves.

Let R' be the tree constructed form R by removing all the leaves.

Since the terminals appear only at the leaves, the claim is equivalent to saying that R' has an ADPIL.

The claim is proved by showing, by contradiction, that there is a root-to-leaf path in R' with at least $m + 1$ nodes,

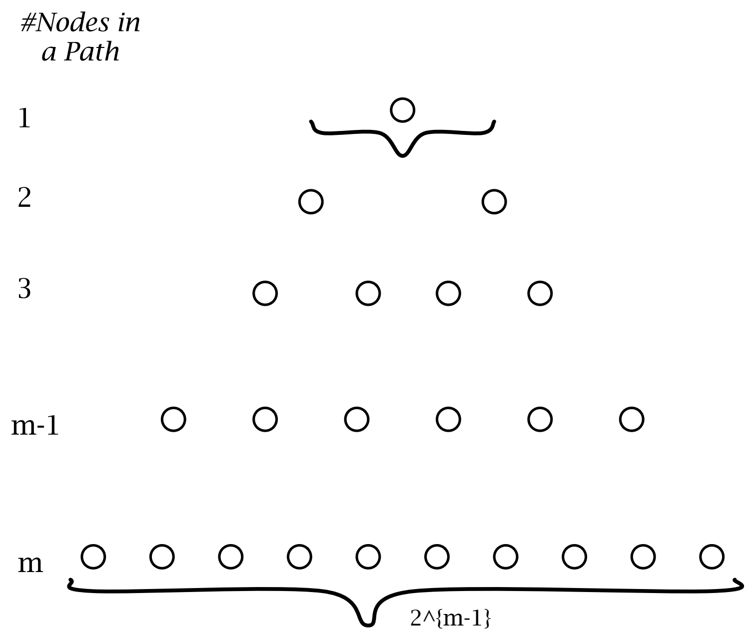
Proof (cont'd)

Assume, by way of contradiction, that every root-to-leaf path of R' has at most m nodes,

Proof (cont'd)

Assume, by way of contradiction, that every root-to-leaf path of R' has at most m nodes,

Then the number of branches in any such path is at most $m-1$. Since R' is a binary tree, R' has at most 2^{m-1} leaves.



Proof (cont'd)

However, the number of leaves of R' is greater than 2^{m-1}. Thus, there is a root-to-leaf path, say π, in R' having length at least $m+1$.

Then, by the pigeonhole principle, an ADPIL appears on π.

Proof of Claim

Proof of Pumping Lemma (cont'd)

Using the following algorithm to find an ADPIL (r, s) farthest from the root of T.

1. Set u to the root of T.
2. Execute the following loop:
   - If the left child of u has an ADPIL, set u to the left child of u.
   - Otherwise, if the right child of u has an ADPIL, set u to the right child of u.
   - Otherwise, quit the loop.
3. Set $r = u$ and s to the leftmost node with the same label as r.

Proof of Pumping Lemma (cont'd)

Using the following algorithm to find an ADPIL (r, s) farthest from the root of T.

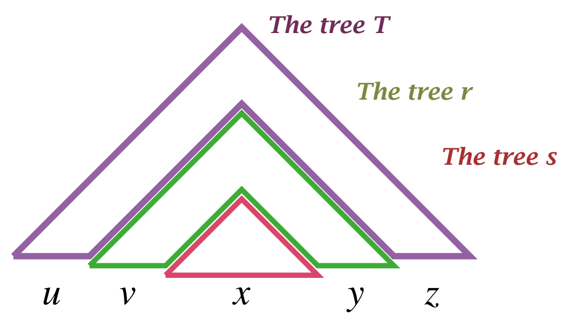
1. Set u to the root of T.
2. Execute the following loop:
   - If the left child of u has an ADPIL, set u to the left child of u.
   - Otherwise, if the right child of u has an ADPIL, set u to the right child of u.
   - Otherwise, quit the loop.
3. Set $r = u$ and s to the leftmost node with the same label as r.

The children of r have no ADPILs. Thus, both children have at most 2^{m-1} leaves and so r has at most $p = 2^m$ leaves.

Proof of Pumping Lemma (cont'd)

Let x be the string at the leaf-level of the subtree rooted at s. Similarly, let vxy be the one for r, where v and y are those to the left and to the right of x, respectively.

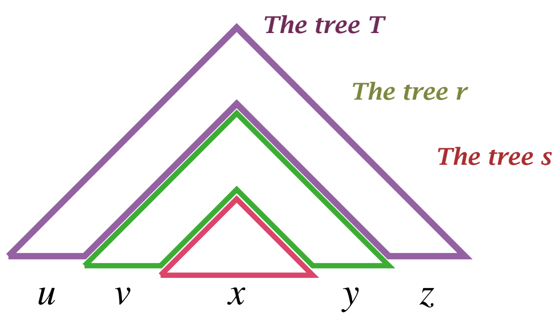
Let u be the string produced to the left of r and z to the right of s.



Proof of Pumping Lemma (cont'd)

Let x be the string at the leaf-level of the subtree rooted at s. Similarly, let vxy be the one for r, where v and y are those to the left and to the right of x, respectively.

Let u be the string produced to the left of r and z to the right of s.

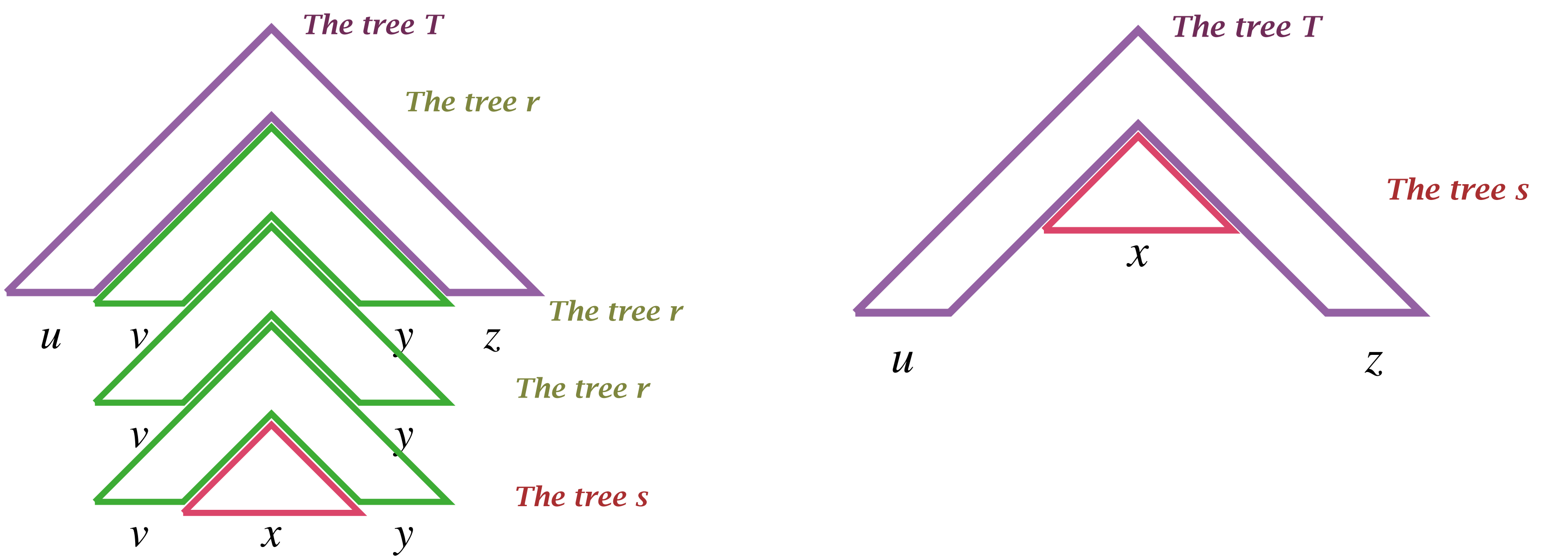


Then $|vxy| \leq p$.

Also, since s is a descendant of r and G has no ϵ-production except for $S \to \epsilon$, x is a proper substring of vxy. Thus, $|vy| \geq 1$.

Proof of Pumping Lemma (cont'd)

Since both r and s have the same label, they are swappable. So, for every $i \geq 0$, $uv^ixy^iz \in L$.



Example 1

$A = \{0^n 1^n 2^n \mid n \geq 0\}$ is not context free.

Example 1

$A = \{0^n 1^n 2^n \mid n \geq 0\}$ is not context free.

Proof Assume, to the contrary, that A is context free. By Pumping Lemma there exists a constant p such that every $w \in A$ of length $\geq p$ is divided into $w = uvxyz$ such that $|vxy| \leq p$, $|vy| \geq 1$, and for every $i \geq 0$, $uv^ixy^iz \in A$.

Let $w = 0^p1^p2^p$. Since $|vxy| \leq p$, vxy is either in 0^*1^* or in 1^*2^*. This means that uv^2xy^2z cannot have the same number of 0s, 1s, as 2s. ∎

Illustrating Conversation

Illustrating Conversation

Illustrating Conversation

Illustrating Conversation

Illustrating Conversation

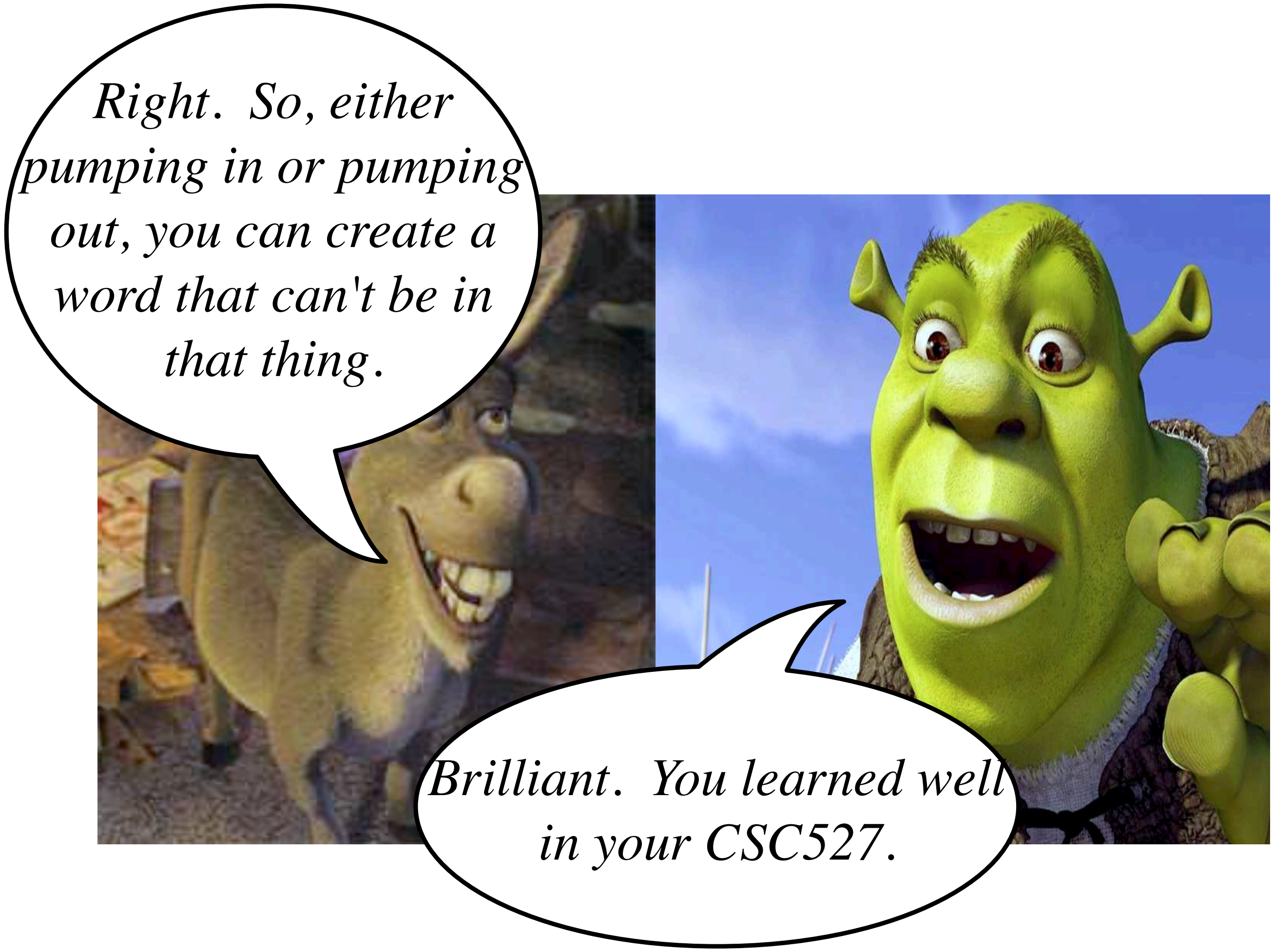
Illustrating Conversation

Illustrating Conversation

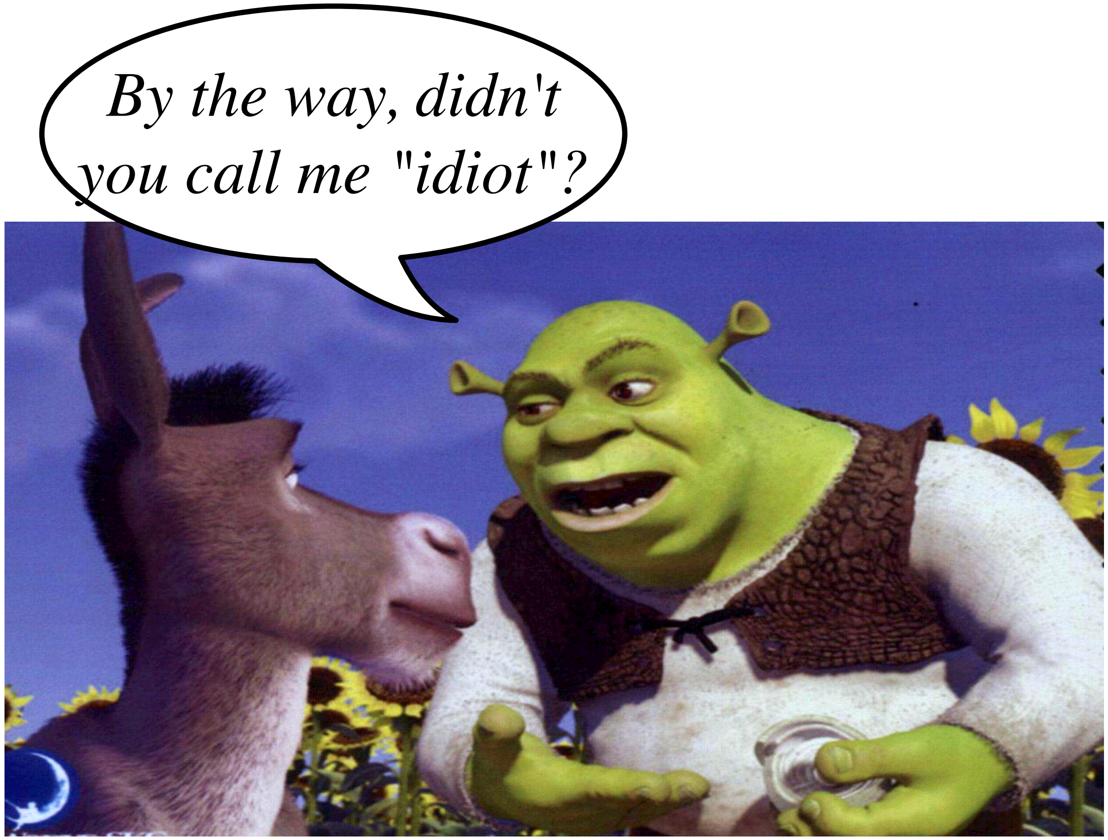
Illustrating Conversation

Illustrating Conversation

Illustrating Conversation



Illustrating Conversation



Example 2

$B = \{a\#b\#c \mid a, b$ and c are binary numbers such that $a + b = c\}$ is not context free.

Example 2

$B = \{a\#b\#c \mid a, b$ and c are binary numbers such that $a + b = c\}$ is not context free.

Proof Assume, to the contrary, that B is context free. Let p be the constant from Pumping Lemma for B. Let $w = 10^p\#10^p\#10^{p+1}$, where $a = b = 2^p$ and $c = 2^{p+1}$. Let $uvxyz$ be the decomposition of w as in the lemma.

Example 2

$B = \{a\#b\#c \mid a, b$ and c are binary numbers such that $a + b = c\}$ is not context free.

Proof Assume, to the contrary, that B is context free. Let p be the constant from Pumping Lemma for B. Let $w = 10^p\#10^p\#10^{p+1}$, where $a = b = 2^p$ and $c = 2^{p+1}$. Let $uvxyz$ be the decomposition of w as in the lemma.

Since each member of B has exactly two $\#$'s, neither v nor y contain a $\sharp$. So, v must be a substring of a, a substring of b, or a substring of c. The same holds for y.

Example 2 (cont'd)

Since the equality $a + b = c$ must be maintained during pumping and $|vy| \geq 1$, y must be a nonempty substring of c and v must be either a nonempty substring of a or a nonempty substring of b. However, since vxy has length at most p, it must be the case that v is a nonempty substring of b.

Example 2 (cont'd)

Since the equality $a + b = c$ must be maintained during pumping and $|vy| \geq 1$, y must be a nonempty substring of c and v must be either a nonempty substring of a or a nonempty substring of b. However, since vxy has length at most p, it must be the case that v is a nonempty substring of b.

If v contains the first letter of b, then vxy must be a part of b, so v should consist solely of 0's.

Example 2 (cont'd)

Since the equality $a + b = c$ must be maintained during pumping and $|vy| \geq 1$, y must be a nonempty substring of c and v must be either a nonempty substring of a or a nonempty substring of b. However, since vxy has length at most p, it must be the case that v is a nonempty substring of b.

If v contains the first letter of b, then vxy must be a part of b, so v should consist solely of 0's.

Suppose y contains 1, the first letter of c. Then uvz is the form $10^p\#10^q\#0^r$, which clearly is not a member of B.

Example 2 (cont'd)

Since the equality $a + b = c$ must be maintained during pumping and $|vy| \geq 1$, y must be a nonempty substring of c and v must be either a nonempty substring of a or a nonempty substring of b. However, since vxy has length at most p, it must be the case that v is a nonempty substring of b.

If v contains the first letter of b, then vxy must be a part of b, so v should consist solely of 0's.

Suppose y contains 1, the first letter of c. Then uvz is the form $10^p\#10^q\#0^r$, which clearly is not a member of B.

Suppose y does not contain the letter 1. Then y consists solely of 0s. Then $uvvxyyz$ is of the form $10^p\#10^q\#10^r$ such that $q, r > p$, which clearly isn't a member of B. ∎

Example 3

$C = \{ww \mid w \in \{0,1\}^*\}$ is not context free.

Example 3

$C = \{ww \mid w \in \{0,1\}^*\}$ is not context free.

Proof Assume C is context free. Let p the constant from the pumping lemma for C.

Let $w = 0^p 1^p 0^p 1^p$. Then w in C.

Example 3

$C = \{ww \mid w \in \{0,1\}^*\}$ is not context free.

Proof Assume C is context free. Let p the constant from the pumping lemma for C.

Let $w = 0^p 1^p 0^p 1^p$. Then w in C.

Let $w = uvxyz$ be the decomposition of w such that $|vy| > 0$, $|vxy| \leq p$, and for every $i \geq 0$, $uv^i xy^i z \in C$.

Example 3

$C = \{ww \mid w \in \{0,1\}^*\}$ is not context free.

Proof Assume C is context free. Let p the constant from the pumping lemma for C.

Let $w = 0^p 1^p 0^p 1^p$. Then w in C.

Let $w = uvxyz$ be the decomposition of w such that $|vy| > 0$, $|vxy| \le p$, and for every $i \ge 0$, $uv^i xy^i z \in C$.

If v contains a symbol from the first 0^p then y cannot contain one from the second 0^p, so pumping doesn't work.

Example 3

$C = \{ww \mid w \in \{0,1\}^*\}$ is not context free.

Proof Assume C is context free. Let p the constant from the pumping lemma for C.

Let $w = 0^p1^p0^p1^p$. Then w in C.

Let $w = uvxyz$ be the decomposition of w such that $|vy| > 0$, $|vxy| \leq p$, and for every $i \geq 0$, $uv^ixy^iz \in C$.

If v contains a symbol from the first 0^p then y cannot contain one from the second 0^p, so pumping doesn't work.

If v contains only symbols from the first 1^p then y cannot contain one from the second 1^p, so pumping doesn't work.

Example 3

$C = \{ww \mid w \in \{0,1\}^*\}$ is not context free.

Proof Assume C is context free. Let p the constant from the pumping lemma for C.

Let $w = 0^p 1^p 0^p 1^p$. Then w in C.

Let $w = uvxyz$ be the decomposition of w such that $|vy| > 0$, $|vxy| \le p$, and for every $i \ge 0$, $uv^i xy^i z \in C$.

If v contains a symbol from the first 0^p then y cannot contain one from the second 0^p, so pumping doesn't work.

If v contains only symbols from the first 1^p then y cannot contain one from the second 1^p, so pumping doesn't work.

If v contains only symbols from the second $0^p 1^p$ then pumping does not work. ∎

Application

Corollary. **The class of context-free languages is not closed under intersection.**

Proof Let $L_1 = \{0^i 1^j 2^k \mid i = j\}$ and $L_2 = \{0^i 1^j 2^k \mid j = k\}$. Then L_1 and L_2 are both context free. If the class were closed under intersection then $L_1 \cap L_2 = \{0^n 1^n 2^n \mid n \geq 0\}$ would be context free. ∎

Corollary. **The class of context-free languages is not closed under complement.**

Proof We know that the class is closed under union. It the class were closed under complement, then by DeMorgan's Law, it would be closed under intersection. ∎